

Validitas Instrumen

Suandi Sidauruk

(Dosen Prodi. Pendidikan Kimia FKIP Univ. Palangkaraya)

Abstrak: Mengukur hasil belajar salah satu cara untuk mengetahui keberhasilan kegiatan pembelajaran. Oleh karena itu, alat ukur yang baik sangat diperlukan untuk mendapat informasi yang akurat. Validitas adalah salah indikator kualitas alat ukur yang harus dipenuhi. Model tunggal gagal membuktikan validitas suatu konstruk. Artinya, inferensi menggunakan cara induktif atau deduktif secara terpisah memiliki kelemahan. Untuk mengatasi hal ini, ditawarkan penggabungan antara metode kuantitatif dengan metode kualitatif. Artinya, proses validasi suatu konstruk menuntut aneka metode daripada metode khusus, wawasan teoretis, dan perhatian terhadap detail.

Kata kunci: Validitas, analisis faktor, MDS

PENDAHULUAN

Pada *Report of the Task Force on Chemical Education Research* dinyatakan bahwa salah satu fokus penelitian pendidikan adalah menelusuri cara siswa membangun pemahaman siswa tentang konsep-konsep bidang studi dalam mentalnya (*Journal of Chemical Education*, 1994). Oleh karena itu, instrumen penelitian pendidikan seyogyanya berupa tes yang dapat mengungkap pemahaman siswa tentang konsep yang telah dipelajarinya.

Telah banyak bentuk tes yang dikembangkan untuk mengukur pemahaman konsep siswa. Secara umum, bentuk tes ini dikelompokkan menjadi dua berdasarkan cara menskor hasil pekerjaan siswa, yaitu subjektif dan objektif. Tes dikatakan bentuk subjektif karena dimungkinkan terdapat skor yang berbeda dari dua orang atau lebih pemberi skor untuk seorang peserta tes. Tes dikatakan bentuk objektif, jika hasil peskoran sama untuk setiap peserta tes dari dua orang atau lebih pemberi skor.

Kelemahan tes bentuk subjektif adalah faktor pemberi skor berpengaruh terhadap skor peserta tes, kelemahan ini tidak terdapat pada tes bentuk objektif. Pada tes bentuk objektif, skor peserta tes semata-mata ditentukan oleh alat ukurnya. Secara teknis, tes bentuk objektif lebih akurat memberikan informasi terkait dengan skor peserta tes. Kegiatan pemberian skor disebut pengukuran yang bersifat kuantitatif. Jika skor ini dimaknai sebagai baik atau buruk, maka kegiatan ini disebut menilai. Contoh menilai hasil belajar seorang peserta tes adalah siswa mendapat skor 8 dikategorikan baik (menguasai konsep).

Pertanyaan berikutnya adalah apakah skor 8 benar-benar mencerminkan penguasaan konsep yang baik? Jika siswa SD kelas 6 diberi soal matematika kelas 1 SMA lalu mendapat skor 1, apakah siswa ini patut dikategorikan tidak baik penguasaan konsepnya? Pertanyaan-pertanyaan ini terkait dengan indikator kualitas tes, seperti validitas dan reliabilitas. Tulisan ini mencoba membahas tentang validitas tes.

PEMBAHASAN

Pengertian Validitas

Validitas adalah proses menginterpretasikan suatu data dengan cara tertentu (Cronbach, 1971). Oleh karena itu, validitas suatu instrumen bersifat relatif tergantung kepada situasi sosial dan tujuannya. Oleh karena sifat ini dan dalam prakteknya bukti-bukti yang ada selalu tidak lengkap, maka validitas suatu instrumen hanya merupakan usaha untuk mencari kebenaran yang rasional berdasarkan pada bukti yang tersedia. Bukti yang tersedia dapat berupa skor, hasil pengamatan, atau alat instrumen lainnya. Dengan demikian yang divalidasi adalah inferensi yang berasal dari skor tes atau indikator lain dan bukan tes atau alat observasi. Selama ini kegiatan inferensi dalam hal validasi didasarkan oleh paradigma positivistik, yaitu melakukan fisikalisasi terhadap fenomena sosial, konstruk dibangun berdasarkan teori untuk menjelaskan fenomena (*theory then research*). Inferensi menggunakan paradigma fenomenologi (*research then theory*) dalam kegiatan membangun konstruk jarang digunakan.

Bukti validitas suatu konstruk seharusnya dapat dijelaskan oleh berbagai cara baik inferensi secara induktif maupun inferensi secara deduktif. Artinya, metode tunggal untuk membuktikan validitas cenderung gagal.

Bukti Validitas

Teori penelitian (sosial) dan teori pengukuran itu sebenarnya satu dan terpadu. Ilmu-ilmu sosial/pendidikan memberikan perhatian terutama terhadap interaksi antara manusia dan makna yang ada dibalik interaksi tersebut. Interaksi ini bersifat kontekstual dan berdimensi jamak. Pada riset konvensional linier, variabel diisolasi, didefinisikan, dan dikategorikan secara teoritis (bebas konteks dan berdimensi tunggal), kemudian menghubungkan variabel tersebut untuk membuat hipotesa sebelum data dikumpulkan. Agar dapat diukur, variabel (konsep-konsep sosial) perlu dijabarkan dalam bentuk operasional sehingga menjadi sangat matematis dan terukur, seperti psikometrik, ekonometrik, atau sosiometrik.

Penelitian sosial sangat berminat pada objek yang bersifat kompleks dan *elusive*, seperti kebencian, pola asuh, dan lain-lain. Variabel yang dibangun dari objek seperti ini bersifat abstrak dan mustahil untuk dicandra. Agar dapat diukur, maka variabel tersebut perlu dijabarkan dalam bentuk operasional, yaitu menyatakan secara tegas dan rinci isyarat (indikator) yang menunjukkan sifat variabel yang diteliti. Selanjutnya, berdasarkan batasan yang ada pada indikator dapat dibuat *indicants* (butir-butir tes).

Berdasarkan uraian di atas, maka arus dari teori ke butir tes adalah:

Teori → Variabel → Definisi Operasional → Indikator → Indicants

Indicants (isi alat ukur yang mampu mengukur apa yang hendak diukur. Pertanyaan ini menunjuk kepada validitas (kesahihan) instrumen, yaitu: keterwakilan isi dalam instrumen (validitas isi); hubungannya dengan kriteria luar yang telah baku (validitas yang dikaitkan dengan kriteria); seberapa jauh instrumen mengukur sifat atau konstruk yang hendak diukur (validitas konstruk) (Kerlinger, 1990; Zeller, 1990).

Validitas Isi

Validitas isi menunjuk kepada sejauh-mana suatu instrumen mencerminkan isi yang hendak diukur. Oleh karena itu, untuk mendapatkan validitas isi yang baik, maka langkah yang dilakukan adalah:

- Menetapkan definisi konsep yang hendak diukur, sehingga diperoleh gambaran semesta isi konsep yang hendak diukur. Semesta ini masih bersifat teoritis dan terbuka. Operasionalisasi definisi ini dapat dilakukan dengan membuat kisi-kisi.
- Mengembangkan dan memilih item yang mencerminkan cakupan isi konsep. Item dirancang menjadi spesifik dan eksak.
- Hindarkan sedapat mungkin faktor-faktor yang tidak ada hubungannya dengan tujuan pengukuran, misalnya hasil tes kimia hendaknya tidak dipengaruhi oleh kemampuan matematika.
- Gunakan evaluator eksternal (pakar, guru) untuk memeriksa isi tes dan mengevaluasi relevansinya dengan semesta yang telah ditentukan.

Hal yang perlu diperhatikan adalah suatu tes yang memiliki validitas tinggi bagi penyusunnya, belum tentu tinggi bagi pemakai lain yang mungkin merumuskan semesta isinya lain.

Uraian ini menjelaskan bahwa informasi tentang kualitas bukti kualitas isi sangat ditentukan oleh ketepatan pemilihan rater. Pada prakteknya, dan banyak dilakukan oleh para peneliti, penilaian validitas isi dilakukan dengan cara menilai sejauhmana butir soal dapat mengukur tujuan dengan cara memberi skor “0” jika butir soal tidak mengukur tujuan, “1” jika butir soal ragu mengukur tujuan, “2” jika butir soal mengukur tujuan.

Cara ini biasanya cukup menggunakan dua atau tiga *rater*. Informasi bukti validitas isi menggunakan cara ini tentu lebih “sederhana” dibandingkan jika menggunakan *rater* lebih banyak. Roid & Haladyna (1982: 210) memperkenalkan cara menghitung tingkat kesesuaian butir soal–tujuan untuk setiap pasangan butir-tujuan menggunakan rumus:

$$KBT_{(t-1)} = \frac{(T-1)S_1 - S_2 \div S_1}{2(T-1)V}$$

T = jumlah tujuan; S₁ = jumlah skor yang diberikan oleh penilai untuk butir 1- tujuan 1;

S₂ = jumlah skor yang diberikan oleh penilai untuk butir 1 – semua tujuan;

V = jumlah validator.

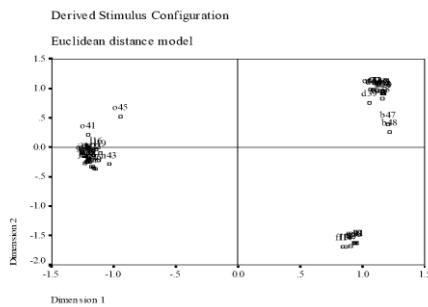
Selanjutnya, informasi bukti validitas isi dapat juga diperoleh dengan cara menghitung tingkat kesejajaran antar butir soal (KBS) yang dinilai oleh *rater*, dalam hal ini guru atau pakar. Sireci & Geisinger (1992) memperkenalkan cara menghitung tingkat kesejajaran butir soal menggunakan teknik skala multidimensi (*multidimensional scaling*, MDS). Konsep MDS didasarkan atas perhitungan terhadap jarak (psikologi) antara dua objek, misal *i* dan *j*, dengan rumus (Kruskal & Wish, 1986: 19):

$$d_{ij} = \sqrt{\sum_{r=1}^R (x_{ir} - x_{jr})^2}$$

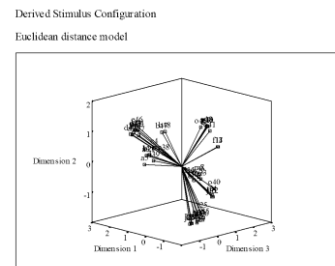
Analisis KBS dapat dilakukan menggunakan program SPSS 10.00 for Windows dengan model ALSCAL (*alternating least-squares scaling*) MDS. Melalui cara ini diperoleh

konfigurasi kesejajaran butir soal serta harga Stress (*square root of a normalized residual sum of square*) dan harga RSQ (*squared correlation*). Tingkat kecocokan yang terbaik diperoleh jika pasangan harga Stress kecil dan harga RSQ besar.

Sidauruk (2005) menggunakan perhitungan dimensi-2, dimensi-3, dimensi-4, dan dimensi-5 untuk menghitung KBS diperoleh harga pasangan Stress-RSQ berturut-turut 0,15-0,92; 0,13-0,92; 0,11-0,93; 0,10-0,93.



Gambar 1. Orientasi butir soal, dimensi-2.



Gambar 2. Orientasi butir soal, dimensi-5.

Secara visual, orientasi butir soal pada solusi dimensi-2 mengelompok menjadi tiga kelompok. Pada solusi dimensi yang lebih tinggi, orientasi butir soal mengalami perubahan dan cenderung mengelompok sesuai dengan tujuan dan ciri butir soal. Ciri butir soal yang menjadi pertimbangan para *rater* dalam menentukan tingkat kesejajaran butir soal adalah soal bentuk tradisional (dapat diselesaikan menggunakan pengetahuan prosedural, *algorithmic-dependent*) atau konseptual (dapat diselesaikan menggunakan pengetahuan deklaratif). Sireci & Geisinger (1992) menyatakan kendala menggunakan dimensi-tinggi adalah kesulitan menerjemahkan hasil perhitungan, sehingga *goodness of fit* tidak cukup digunakan untuk menentukan dimensionalitas.

Validitas Tekait dengan Kriteria

Berbeda dengan validitas isi, validitas yang dikaitkan dengan kriteria menunjuk pada hubungan antara skor suatu instrumen pengukuran dengan suatu variabel (kriteria) luar yang mandiri dan dipercaya dapat mengukur langsung tingkah laku atau ciri-ciri yang diselidiki. Validitas ini lebih memberi tekanan pada kriterianya, bukan pada instrumen itu sendiri. Hal yang terutama diperhatikan adalah apa yang dapat diramalkan oleh instrumen tersebut bukan isi tesnya.

Oleh karena itu, identifikasi kriteria yang dipakai menjadi penting. Ciri-ciri kriteria yang baik adalah: Relevansi, kriteria yang dipilih benar-benar menggambarkan ciri-ciri yang tepat dari perilaku/atribut yang diselidiki; Reliabel, kriteria tersebut harus merupakan ukuran yang ajeg bagi atribut tersebut; Bebas dari bias, pemberian skor pada suatu ukuran kriteria hendaknya tidak dipengaruhi oleh faktor-faktor selain penampilan sebenarnya pada kriteria itu

Berdasarkan waktu kapan data criteria dikumpulkan, validitas ini dibedakan menjadi dua macam, yaitu: validitas prediktif dan validitas konkuren. Validitas prediktif berkenaan dengan korelasi antara skor tes dengan suatu kriteria yang terjadi dikemudian hari. Validitas konkuren berkenaan dengan korelasi antara skor tes dengan suatu ukuran kriteria yang diperoleh pada waktu yang sama atau pada waktu yang hampir bersamaan.

Validitas Konstruk

Validitas ini menunjuk kepada seberapa jauh hasil tes dapat ditafsirkan menurut konstruk tersebut. Konstruk dipakai untuk menunjukkan sesuatu yang tidak dapat diukur secara langsung, tetapi yang dapat menerangkan akibat-akibat yang dapat diamati.

Untuk menetapkan validitas konstruk tidak hanya ada satu cara saja. Bukti-bukti dikumpulkan dari berbagai sumber, termasuk data validitas isi dan validitas yang dikaitkan dengan kriteria. Setiap data yang dapat membantu menafsirkan arti skor tes dianggap sebagai data yang relevan.

Beberapa cara yang digunakan untuk menyelidiki validitas konstruk, antara lain, adalah:

- Korelasi dengan ukuran yang lain. Suatu tes hendaknya mempunyai korelasi yang tinggi dengan ukuran lain yang dianggap sebagai ukuran yang valid bagi konstruk tersebut. Tes TPA hendaknya berkorelasi tinggi dengan tes IQ.
- Analisis faktor. Metoda statistik ini digunakan untuk menyelidiki hubungan antar seperangkat skor tes serta menetapkan jumlah faktor (konstruk) yang diperlukan untuk menerangkan hubungan itu.. Hubungan ini menunjukkan bukan saja tes-tes mana yang mengukur faktor yang sama, tetapi juga seberapa jauh tes-tes tersebut mengukur faktor-faktor tersebut.
- Studi eksperimental. Metoda ini digunakan untuk menguji hipotesa bahwa skor tes akan berubah apabila diberikan beberapa jenis perlakuan eksperimental kepada subjek.
- Perbandingan skor kelompok-kelompok tertentu. Metoda ini digunakan untuk menguji validitas konstruk suatu tes dengan cara mengujikan tes tersebut kepada sekelompok subjek yang telah dikenali memiliki konstruk yang diukur dengan sekelompok subjek lain yang tidak memiliki sifat-sifat konstruk yang diukur.
- Analisis intra tes. Dalam metode ini, peneliti memeriksa tes itu sendiri serta mengumpulkan informasi tentang isi tes tersebut, proses yang digunakan dalam menjawab pertanyaan tes, dan korelasi antara butir-butir tes tersebut.

Kritik terhadap validitas konstruk adalah bila penekanan terhadap validitas ini terlalu banyak, maka dapat mengakibatkan diremehkannya segi praktis dari pengujian dan evaluasi.

PENUTUP

Di muka telah dikemukakan bahwa perkembangan pengukuran sejalan dengan perkembangan penelitian sosial. Selama ini penelitain sosial dikembangkan melalui paradigma positivistik-kuantitatif. Salah satu ciri dari paradigma ini adalah terukur, sehingga tidak dapat dihindarkan fisikalisasi konsep-konsep sosial. Perkembangan teori sosial mengadopsi cara-cara yang digunakan pada IPA. Pandangan terhadap fenomena kehidupan

sosial yang menekankan aspek rasional menjadi bias dan menghasilkan gambaran kehidupan sosial yang artifisial.

Ilmu-ilmu sosial/pendidikan memberikan perhatian terutama terhadap interaksi antara manusia dan makna yang ada dibalik interaksi tersebut. Interaksi ini bersifat kontekstual dan berdimensi jamak (Zeller, 1990). Oleh karena itu, konsep (konstruk) hanya dapat didefinisikan secara objektif bila informasi berasal dari realita itu sendiri. Pendefinisian konstruk seperti ini dilakukan dengan cara inferensi deduktif. Kritik terhadap cara ini adalah menentukan kriteria untuk mengetahui akan kebenaran berdasarkan hasil inferensi.

Beberapa pakar terbukti telah gagal membuktikan validitas suatu konstruk menggunakan metode tunggal. Artinya, inferensi menggunakan cara induktif atau deduktif secara terpisah memiliki kelemahan. Untuk mengatasi hal ini, ditawarkan penggabungan antara metode kuantitatif dengan metode kualitatif. Artinya, proses validasi suatu konstruk menuntut aneka metode daripada metode khusus, wawasan teoretis, dan perhatian terhadap detail.

DAFTAR RUJUKAN

- Allen, M.J. & Yen, W.M. 1979. *Introduction to Measurement Theory*. Monterey: Brooks/Cole Publ. Co.
- Cronbach, L.J. 1971. Test Validation. Dalam Robert L. Thorndike (Ed), *Educational Measurement (2nd)* (443-507). American Council on Education. Washington D.C.
- Guilford, J. P. 1954. *Psychometric Methods*. Tokyo: McGraw-Hill Book Co, Inc.
- Kruskal, J. B., & Wish, M. 1986. *Multidimensional Scaling*. London. Sage Publications
- Roid, G. H. & Haladyna, T. M. 1982. *A Technology for Test-item Writing*. Orlando: Academic Press. Inc.
- Sidauruk, S. 2005. *Miskonsepsi Stoikiometri Siswa SMA* (Disertasi). Yogyakarta. PPs UNY.
- Sireci, S.G., & Geisinger, K.F. 1992. Analyzing Test Content Using Cluster Analysis and Multidimensional Scaling. *Applied Psychological Measurement*, 16(1), 17-31.
- Sireci, S.G., & Geisinger, K.F. 1995. Using Subject-Matter Experts to Assess Content Representation: An MDS Analysis. *Applied Psychological Measurement*, 19(3), 241-255.
- Zeller, R.A. 1990. Validity. Dalam John P. Keeves (Ed). *Educational Research, Methodology, and Measurement: an International Handbook*. New York: Pergamon
- Kerlinger, F.N. 1990. *Asas-Asas Penelitian Behavioral* (diterjemahkan oleh: Landung R. Simatupang). Yogyakarta: Gadjah Mada University Press.